

Search for single-gene predictors of cisplatin response

November 15, 2008

1 Load and prepare data

1.1 Load add-on packages

```
> library(affy)
> library(hgu133plus2.db)
> library(aronmisc)
```

1.2 Load the RMA-processed expression values

```
> load("~/aron/sbge_cancer/data/exprs/platinum.rma.RData")
```

1.3 Remove outliers and non-cisplatin-treated samples

```
> ok.sample <- !is.na(platinum.rma$Miller)
> ok.sample[platinum.rma$Chip %in% c("P9", "P20")] <- FALSE
> plat.rma <- platinum.rma[, ok.sample]
> dim(exprs(plat.rma))
```

```
[1] 54675    24
```

1.4 Define the response groups

We are interested in the Miller-Payne response score. Note: MP score = 0 means "progression". We are also interested in the distinction between two response groups: Miller-Payne grades 0-2, and M-P grades 3-5.

```
> MP.score <- plat.rma$Miller.Payne.Grade
> binary.response <- factor(ifelse(MP.score > 2, "MP345", "MP012"))
> table(binary.response, MP.score)
```

	MP.score					
binary.response	0	1	2	3	4	5
MP012	3	5	4	0	0	0
MP345	0	0	0	4	4	4

1.5 Miscellaneous

We will need gene names and symbols for each probe set.

```
> symb <- sapply(mget(featureNames(plat.rma), hgu133plus2SYMBOL),
+   function(x) x[1])
> gene <- sapply(mget(featureNames(plat.rma), hgu133plus2GENENAME),
+   function(x) x[1])
```

Here is a function to calculate a P value for a Pearson correlation coefficient:

```
> pearson.p <- function(r, n, alternative = "two.sided") {
+   df <- n - 2
+   t.stat <- sqrt(df) * r/sqrt(1 - r^2)
+   p <- pt(t.stat, df)
+   switch(alternative, less = p, greater = 1 - p, two.sided = 2 *
+     pmin(p, 1 - p))
+ }
```

2 Find genes correlated with Miller-Payne response score

2.1 Calculate Pearson correlation coefficient for each probe set

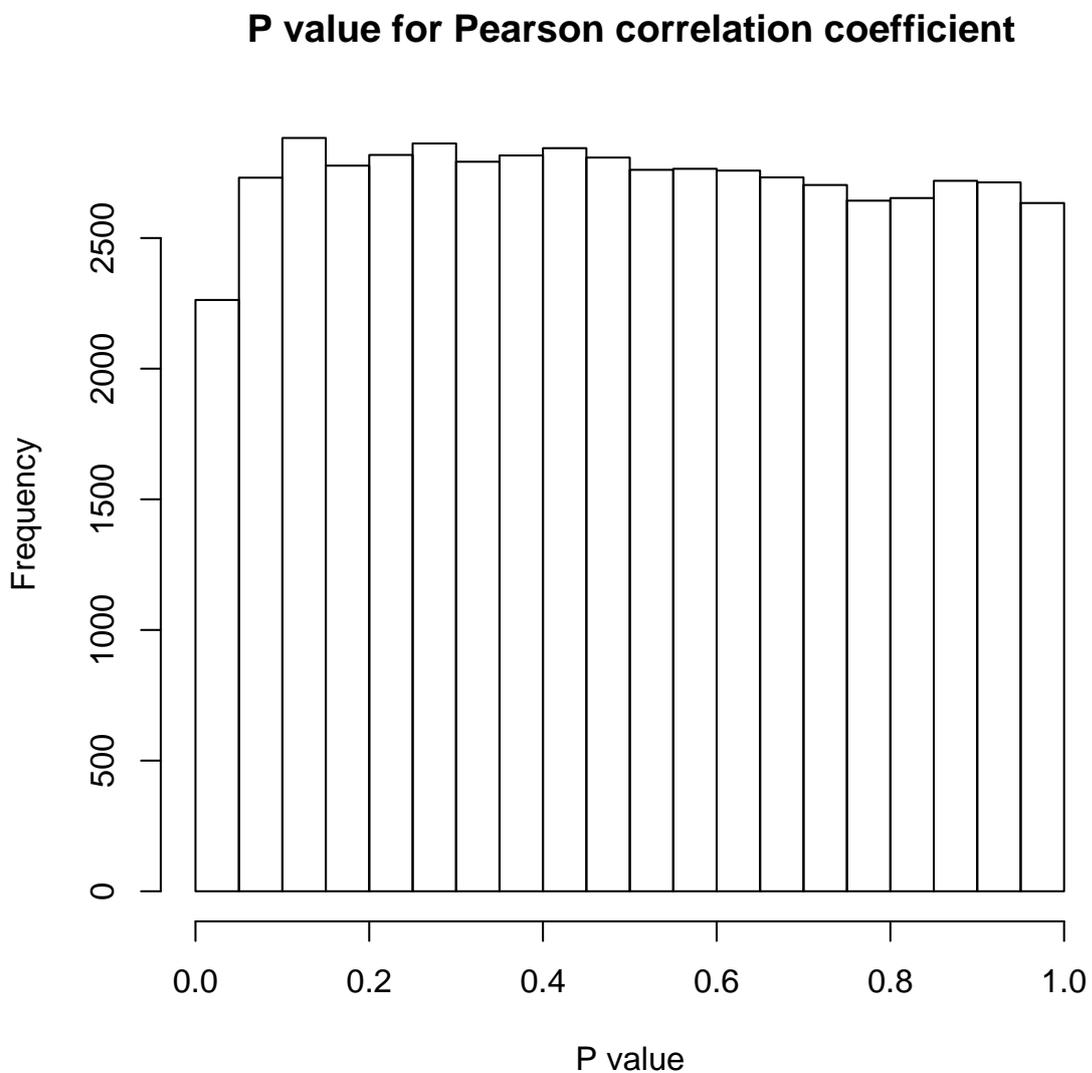
```
> corr <- cor(t(exprs(plat.rma)), MP.score)[, 1]
```

Calculate P values for the correlation coefficients:

```
> corr.p <- pearson.p(corr, n = 24)
```

2.2 Visualize the distribution of P values

```
> hist(corr.p, xlab = "P value", main = "P value for Pearson correlation coefficient")
```

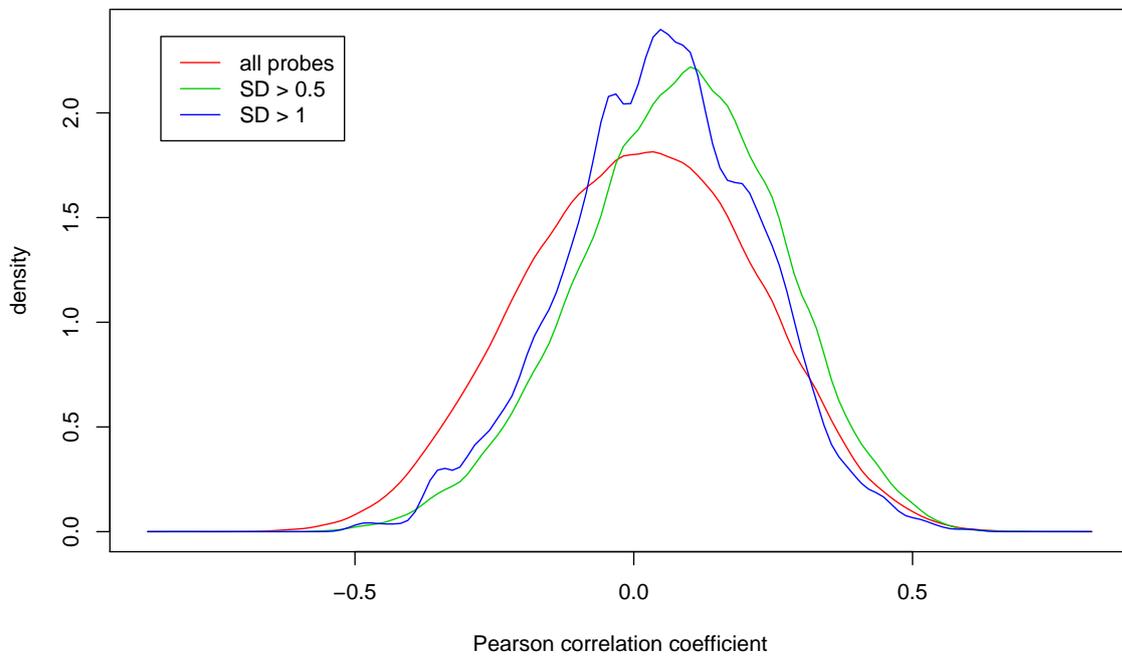


2.3 Filter out those measurements without substantial changes

```
> s <- apply(exprs(plat.rma), 1, sd)
> table(s > 0.5)
```

```
FALSE TRUE
43494 11181
```

```
> corr.bySD <- list("all probes" = corr, "SD > 0.5" = corr[s >
+ 0.5], "SD > 1" = corr[s > 1])
> plot(mdensity(corr.bySD), col = 2:4, xlab = "Pearson correlation coefficient")
> legend("topleft", legend = names(corr.bySD), col = 2:4, lty = 1,
+ inset = 0.05)
```



2.4 List top probe sets

Select potentially interesting probe sets: those with a P value below 0.01, and a standard deviation above 0.5.

```
> is.topcor <- corr.p < 0.01 & s > 0.5
> topcortable <- data.frame(symbol = symb[is.topcor], corr = corr[is.topcor],
+   SD = s[is.topcor], p = corr.p[is.topcor], gene.name = gene[is.topcor])
> topcortable[order(topcortable$corr), ]
```

	symbol	corr	SD	p
218773_s_at	MSRB2	-0.5600596	0.6576005	0.0044253073
201170_s_at	BHLHB2	-0.5433392	0.5462046	0.0060705630
201963_at	ACSL1	-0.5246001	0.9347005	0.0084936698
201636_at	FXR1	0.5153064	0.6511447	0.0099652666
201565_s_at	ID2	0.5191747	0.8157628	0.0093290179
223334_at	TMEM126A	0.5202369	0.6182935	0.0091603190
202095_s_at	BIRC5	0.5234891	0.8456738	0.0086594694
202313_at	PPP2R2A	0.5239942	0.6518257	0.0085837703
238026_at	RPL35A	0.5240831	0.7648467	0.0085705018
224675_at	MESDC2	0.5274197	0.5677276	0.0080846985
225098_at	ABI2	0.5277410	0.6678753	0.0080391526
212952_at	CALR	0.5284795	0.5557848	0.0079352787
230005_at	SVIP	0.5329421	1.0247749	0.0073310458
225358_at	DNAJC19	0.5329652	0.6487480	0.0073280197
202550_s_at	VAPB	0.5340002	0.5138295	0.0071935522
226880_at	NUCKS1	0.5340370	0.6428383	0.0071888084
224568_x_at	MALAT1	0.5359033	0.7861830	0.0069516628
204026_s_at	ZWINT	0.5372117	0.8741038	0.0067893279
202028_s_at	RPL38	0.5376327	0.6161097	0.0067377696
201051_at	ANP32A	0.5381672	0.7815489	0.0066727892
201637_s_at	FXR1	0.5423241	0.6683103	0.0061850303
202370_s_at	CBFB	0.5434405	0.6266835	0.0060592332
225679_at	NAT12	0.5474180	0.5364236	0.0056283245
221521_s_at	GINS2	0.5494674	0.6583849	0.0054165176
237180_at	PSME4	0.5502225	0.7201382	0.0053401852
223062_s_at	PSAT1	0.5505508	1.3998033	0.0053072751
224607_s_at	SRP68	0.5523267	0.5915462	0.0051322040
218374_s_at	C12orf4	0.5542205	0.6213348	0.0049508788
227719_at	<NA>	0.5561575	0.5414437	0.0047710124
1554080_at	RQCD1	0.5585944	0.5992970	0.0045525577
212905_at	CSTF2T	0.5616656	0.9664655	0.0042892835
226784_at	TWISTNB	0.5657722	0.6031918	0.0039573771
203341_at	CEBPZ	0.5712300	0.6339654	0.0035499910
222613_at	C12orf4	0.5722430	0.6995576	0.0034784208
218011_at	UBL5	0.5839931	0.5313749	0.0027334056
231735_s_at	PR01073	0.5992404	1.0056062	0.0019720300
236814_at	MDM4	0.6099862	0.7284398	0.0015513050

226574_at	PSPC1	0.6185723	0.5607219	0.0012727218	
223577_x_at	MALAT1	0.6226103	0.7093162	0.0011572993	
205070_at	ING3	0.6859876	0.5073412	0.0002151604	
					gene.name
218773_s_at					methionine sulfoxide reductase B2
201170_s_at					basic helix-loop-helix domain containing, class B, 2
201963_at					acyl-CoA synthetase long-chain family member 1
201636_at					fragile X mental retardation, autosomal homolog 1
201565_s_at					inhibitor of DNA binding 2, dominant negative helix-loop-helix protein
223334_at					transmembrane protein 126A
202095_s_at					baculoviral IAP repeat-containing 5 (survivin)
202313_at					protein phosphatase 2 (formerly 2A), regulatory subunit B, alpha isoform
238026_at					ribosomal protein L35a
224675_at					mesoderm development candidate 2
225098_at					abl interactor 2
212952_at					calreticulin
230005_at					small VCP/p97-interacting protein
225358_at					DnaJ (Hsp40) homolog, subfamily C, member 19
202550_s_at					VAMP (vesicle-associated membrane protein)-associated protein B and C
226880_at					nuclear casein kinase and cyclin-dependent kinase substrate 1
224568_x_at					metastasis associated lung adenocarcinoma transcript 1 (non-protein coding)
204026_s_at					ZW10 interactor
202028_s_at					ribosomal protein L38
201051_at					acidic (leucine-rich) nuclear phosphoprotein 32 family, member A
201637_s_at					fragile X mental retardation, autosomal homolog 1
202370_s_at					core-binding factor, beta subunit
225679_at					N-acetyltransferase 12
221521_s_at					GINS complex subunit 2 (Psf2 homolog)
237180_at					proteasome (prosome, macropain) activator subunit 4
223062_s_at					phosphoserine aminotransferase 1
224607_s_at					signal recognition particle 68kDa
218374_s_at					chromosome 12 open reading frame 4
227719_at					<NA>
1554080_at					RCD1 required for cell differentiation1 homolog (S. pombe)
212905_at					cleavage stimulation factor, 3' pre-RNA, subunit 2, 64kDa, tau variant
226784_at					TWIST neighbor
203341_at					CCAAT/enhancer binding protein zeta
222613_at					chromosome 12 open reading frame 4
218011_at					ubiquitin-like 5
231735_s_at					PRO1073 protein
236814_at					Mdm4, transformed 3T3 cell double minute 4, p53 binding protein (mouse)
226574_at					paraspeckle component 1
223577_x_at					metastasis associated lung adenocarcinoma transcript 1 (non-protein coding)
205070_at					inhibitor of growth family, member 3

2.5 Generate figures for a few top probe sets

```
> top9 <- topcortable[order(abs(topcortable$corr), decreasing = TRUE)[1:9],
+ ]
> top9
```

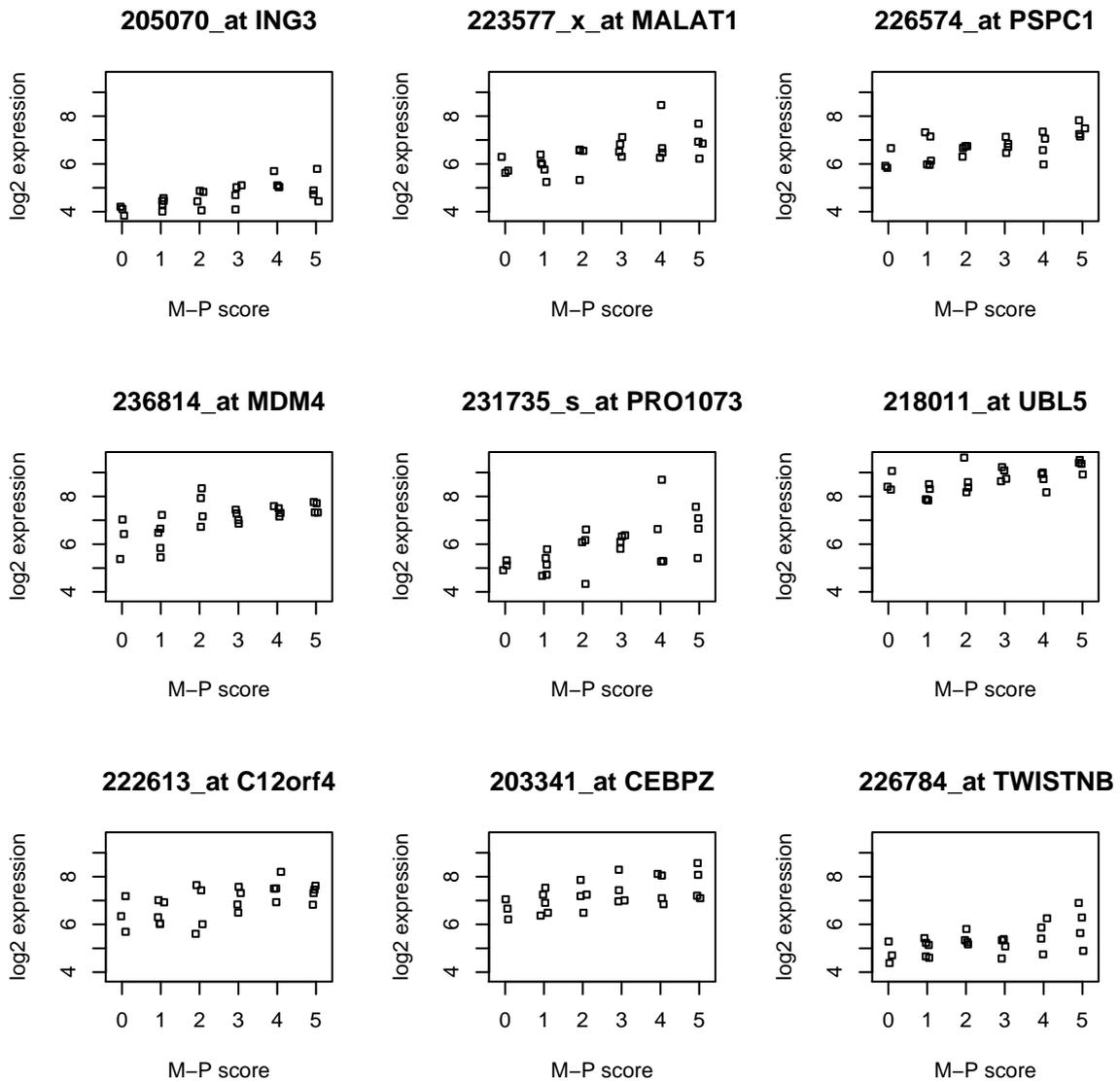
	symbol	corr	SD	p
205070_at	ING3	0.6859876	0.5073412	0.0002151604
223577_x_at	MALAT1	0.6226103	0.7093162	0.0011572993
226574_at	PSPC1	0.6185723	0.5607219	0.0012727218
236814_at	MDM4	0.6099862	0.7284398	0.0015513050
231735_s_at	PR01073	0.5992404	1.0056062	0.0019720300
218011_at	UBL5	0.5839931	0.5313749	0.0027334056
222613_at	C12orf4	0.5722430	0.6995576	0.0034784208
203341_at	CEBPZ	0.5712300	0.6339654	0.0035499910
226784_at	TWISTNB	0.5657722	0.6031918	0.0039573771

	gene.name
205070_at	inhibitor of growth family, member 3
223577_x_at	metastasis associated lung adenocarcinoma transcript 1 (non-protein coding)
226574_at	paraspeckle component 1
236814_at	Mdm4, transformed 3T3 cell double minute 4, p53 binding protein (mouse)
231735_s_at	PR01073 protein
218011_at	ubiquitin-like 5
222613_at	chromosome 12 open reading frame 4
203341_at	CCAAT/enhancer binding protein zeta
226784_at	TWIST neighbor

```

> par(mfrow = c(3, 3))
> ylim <- range(exprs(plat.rma)[rownames(top9), ])
> for (i in rownames(top9)) {
+   stripchart(exprs(plat.rma)[i, ] ~ MP.score, ylim = ylim,
+             method = "jitter", vertical = TRUE, xlab = "M-P score",
+             ylab = "log2 expression")
+   title(main = paste(i, symb[i]))
+ }

```



3 sessionInfo

The results in this file are generated using the following packages:

```
> sessionInfo()
```

```
R version 2.7.2 (2008-08-25)
```

```
i386-apple-darwin8.11.1
```

```
locale:
```

```
C
```

```
attached base packages:
```

```
[1] splines    tools      stats      graphics  grDevices  utils      datasets
```

```
[8] methods   base
```

```
other attached packages:
```

```
[1] aronmisc_0.1.8      survival_2.34-1     squash_0.1.7
```

```
[4] hgu133plus2.db_2.2.0 AnnotationDbi_1.2.2 RSQLite_0.6-9
```

```
[7] DBI_0.2-4           affy_1.18.2         preprocessCore_1.2.0
```

```
[10] affyio_1.8.0        Biobase_2.0.1
```

```
> system("uname -a", intern = TRUE)
```

```
[1] "Darwin eklund.local 9.5.0 Darwin Kernel Version 9.5.0: Wed Sep  3 11:29:43 PDT 2008;
```